AFRL-SA-WP-TR-2016-0008

Operational Assessment of Color Vision

**Steve Wright, O.D.; James Gaska, Ph.D.;
Marc Winterbottom, Ph.D.; Darrell Rousse, O.D.;
Steven Hadley, M.D.; Dan LaMothe, M.D.**

June 2016

**Final Report
for October 2012 to June 2014**

**Air Force Research Laboratory
711th Human Performance Wing
U.S. Air Force School of Aerospace Medicine
Aerospace Medicine Department
Ophthalmology Branch
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-SA-WP-TR-2016-0008 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNATURE//	//SIGNATURE//
COL JOHN P. LYNCH	COL PATRICK R. STORMS
Chief, Aerospace Consultation Service Div	Chair, Aerospace Medicine Department

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
20 Jun 2016	Final Technical Report	October 2012 – June 2014

4. TITLE AND SUBTITLE

Operational Assessment of Color Vision

5a. CONTRACT NUMBER
FA8650-11-D-6233 / FA8650-10-D-6056

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steve Wright, James Gaska, Marc Winterbottom, Darrell Rousse, Steven Hadley, Dan LaMothe

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
USAF School of Aerospace Medicine
Aerospace Medicine Department
Ophthalmology Branch
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-SA-WP-TR-2016-0008

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES
Cleared, 88PA, Case # 2016-3503, 15 Jul 2016.

14. ABSTRACT

Normal perception of color vision is thought to be an important attribute for many occupations. Several mishaps in the transportation industry have been blamed on color deficiencies. Historically, color vision status was assessed using color matching or color naming tests or pseudo-isochromatic plates. More recently, however, computer-based automated tests have been developed to both improve testing sensitivity and offer the ability to quantify levels of color deficiencies. Color vision testing was performed on 50 color normal and 50 color abnormal subjects using three commercially available computer-based color tests as well as an anomaloscope. These findings were related to a color sorting task that represented an operationally relevant task for U.S. Air Force aviators. The computer-based tests proved to be highly sensitive in identifying color vision deficiencies, in some cases more sensitive than the anomaloscope. Overall, there was a trend of worsening performance on the operational task with increasing levels of color deficiency. While this was a statistically significant finding, the magnitude of the effect was low, and many color abnormal subjects, especially mildly and moderately deficient subjects, performed within normal levels. We recommend these findings be interpreted with caution, as the operational task used colors with very specific chromaticities and with high contrast under optimal conditions. True operational environments are far more austere with variables that go beyond those evaluated in this study.

15. SUBJECT TERMS
Color vision, aviation, cone contrast test, Colour Assessment & Diagnosis, color Dx, OBVA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Steve Wright
U	U	U	SAR	29	19b. TELEPHONE NUMBER *(include area code)*

This page intentionally left blank.

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGMENTS

This work was supported in part by U.S. Air Force Contract FA8650-05-D-6502 to Link Simulation and Training (a division of L3 Communications Corp.) and was funded by the 711th Human Performance Wing, U.S. Air Force School of Aerospace Medicine.

The authors would like to thank several members of the Operational Based Vision Assessment (OBVA) Integrated Product Team: Dr. Jeff Hovis (University of Waterloo), Dr. Frank Kooi (The Netherlands Organisation), Dr. (Lt Col) Jay Flottmann, Dr. (Col) Lex Brown, Lt Col Robert Forino, and Mr. John White, who all contributed to the development of the OBVA Laboratory and identification of operationally relevant flight tasks, including the color situation awareness (SA) display task described here. In particular, we would like to thank Mr. John White, a very experienced F-16 pilot, who designed the color SA display snapshots for this research; Mr. Chi Feng Tai, who developed the MatLab software to create the test plates as well as the software running the experiment; Mr. Scott Humphrey, who supported subject recruitment and data collection; and Mr. Jared Haynes, who provided statistical expertise.

This page intentionally left blank.

1.0 SUMMARY

Normal perception of color vision is thought to be an important attribute for many occupations. Several mishaps in the transportation industry have been blamed on color deficiencies. Historically, color vision status was assessed using color matching or color naming tests or pseudo-isochromatic plates. More recently, however, computer-based automated tests have been developed to both improve testing sensitivity and offer the ability to quantify levels of color deficiencies. Color vision testing was performed on 50 color normal and 50 color abnormal subjects using three commercially available computer-based color tests as well as an anomaloscope. These findings were related to a color sorting task that represented an operationally relevant task for U.S. Air Force aviators. The computer-based tests proved to be highly sensitive in identifying color vision deficiencies, in some cases more sensitive than the anomaloscope. Overall, there was a trend of worsening performance on the operational task with increasing levels of color deficiency. While this was a statistically significant finding, the magnitude of the effect was low, and many color abnormal subjects, especially mildly and moderately deficient subjects, performed within normal levels. We recommend these findings be interpreted with caution, as the operational task used colors with very specific chromaticities and with high contrast under optimal conditions. True operational environments are far more austere with variables that go beyond those evaluated in this study.

2.0 INTRODUCTION

Determination of color vision status has been a longstanding topic of interest for the transportation industry. While many other occupations maintain some level of color vision standards, it is of particular importance in transportation, as colored symbologies are frequently used to aid or direct critical activities such as aircraft landing approaches or railroad right-of-way designations. Misinterpretation of these signals due to anomalous color perception may lead to mishaps, resulting in fatalities and/or significant financial loss.

Color vision deficiency was identified as the primary factor in a 1996 New Jersey railroad collision that killed 3, injured 158, and left $3.3M in damage. In this event, the train engineer neglected to report a medical condition that precipitated an acquired color deficiency. This was compounded by the fact that the medical examiner did not follow proper color vision testing protocol and the condition went undiagnosed. Consequently, the engineer misinterpreted a red stop signal, failed to yield, and struck a commuter train [1]. Defective color vision was similarly cited as a contributing factor in a railroad collision in Oklahoma in 2012 [2]. The engineer suffered from an acquired color vision loss as well as overall reduced visual acuity secondary to multiple ophthalmic pathologies including glaucoma, cataracts, cystoid macular edema, and epi-retinal membranes. Despite his complaints that he had difficulty seeing signals and that he did not meet either color vision or visual acuity standards, the engineer maintained medical certification. The National Transportation Safety Board determined the probable cause of the accident was due to the engineer's inability to see and correctly interpret a colored signal. This event resulted in three fatalities and $14.8M in damage. In 2002, the cause of a Fed Ex 727 crash in Tallahassee, FL, was blamed, in part, on the first officer's inability to properly interpret the precision approach path indicator lights due to a congenital color deficiency [3]. During a visual approach at night, the plane descended below glide slope and landed short of the runway, injuring all three crew members with total loss of the plane. As with most mishaps, there was

not one singular cause cited; however, defective color vision and misinterpretation of the landing lights were cited as contributing to the event.

The origin of occupational color vision screening is somewhat clouded in history. Wilson [4] has suggested that the first color vision standards for railway workers were implemented around 1853 by the Great Northern Railway Company in England, more than a quarter of a century before visual acuity standards were adopted [5]. However, these standards were not established on an industry-wide scale, and there are doubts whether the standards were effectively administered. Reasons for establishing these standards are not abundantly clear, as there was no sentinel event that could be identified as a basis for the change. Vingrys [6] proposed they likely arose due to (1) the demand for safe and reliable methods of travel, (2) the increased use of colored signals in the transportation industry, and (3) the recognition that color vision deficiencies were relatively common and improper recognition of colored signals could result in a mishap. The landmark event that most authorities cite as the true origin of occupational screening for color vision was a fatal railroad accident in Lagerlunda, Sweden, in 1875, described in detail by Mollon [7]. Ironically, there is no conclusive evidence that color vision played a role in the mishap at all. The two individuals who were hypothesized to be color deficient perished in the accident and could not be tested, leaving the authors to conclude that "without doubt the Lagerlunda accident had a central role in the introduction of screening for color deficiency by railroads throughout the world, but it is less certain that color deficiency had a central role in the Lagerlunda accident" [7].

The U.S. military adopted color vision and visual acuity standards for the U.S. Navy and Merchant Marine through House Resolution 135 of the Forty-Seventh Congress 1881-1883 [5], although the specific methods of testing were not stated. Early color vision testing methods employed by the U.S. Army Air Corps consisted of the Jennings' self-recording test [8], with Stilling's pseudo-isochromatic plates and the William's lantern as secondary tests [9]. Over the next five decades, various forms of pseudo-isochromatic plates were utilized by the Army Air Corps (now U.S. Air Force) and were nicely summarized by Tredici et al. in 1972 [10]. In recent years, these tests have fallen under scrutiny and have been described as having "antiquated roots" in relation to the changes in aviation platforms and visual demands on aviators in the 21st century [11]. Furthermore, traditional color vision screening methods have numerous, well-known deficiencies including (1) reduced sensitivity and specificity [12,13], (2) inability to fully randomize the presentation, (3) propensity for technician error and bias, 4) requirement for specialty lighting to ensure test validity, and 5) test plates fade over time.

In the last several decades, improvements to computer-generated display systems have facilitated the development of computer-based, automated tests of color vision [14,15]. The United Kingdom's Civil Aviation Authority now recognizes the Colour Assessment & Diagnosis (CAD) as an authorized color vision screening device for civilian pilots [16]. Similarly, the U.S. Navy is currently evaluating the Waggoner Computerized Color Vision Test (WCCVT) for screening aircrew members [17]. In 2011, the U.S. Air Force transitioned to computer-based testing, with the Rabin Cone Contrast Test (RCCT) designated as the sole approved device for assessment of color vision for aircrew personnel and applicants to aircrew positions [18]. These devices address the deficiencies of traditional color tests mentioned above and additionally quantify the severity of the color deficiency. This was not possible with some of the older tests such as the anomaloscope or lantern tests.

The purpose of this study was to compare the clinical screening performance (sensitivity and specificity) of the RCCT relative to the CAD and WCCVT as well as to the anomaloscope,

which is widely considered the gold standard for assessing color vision status. The second goal of the study was to relate clinical measurements of color vision to performance on a scenario simulating an operational color vision task.

3.0 METHODS

3.1 Participants

This prospective study was approved by the Wright-Patterson Air Force Base Institutional Review Board (IRB # FWR20120220H). All subjects offered consent prior to participation and were free to withdraw at any point during the study. Study participants consisted of 50 individuals with normal color vision and 50 individuals with varying degrees of a congenital deutan or protan color deficiency. All subjects participated in the clinical screening tasks, although two could not be evaluated on the operational task due to time constraints. Inclusion criteria consisted of participant age between 18 and 50 years, uncorrected or corrected visual acuity of at least 20/20 in each eye, and the absence of ocular pathology or medication use that could impair color vision. Table 1 describes the demographics of the subjects evaluated. There was a marked gender bias in the normal subject pool due to the fact that this population was composed primarily of pilot applicants, a traditionally male-dominated field. Additionally, color deficient subjects showed a larger bias due to the fact that congenital color deficiencies are x-linked recessive traits, thus predisposing men to the condition.

Table 1. Subject Demographics

Group	Mean age (range) (yr)	Gender (Male/Female)
Color Normal	29.3 (18-50)	35 / 15
Deutan	32.3 (19-50)	28 / 4
Protan	31.2 (18-50)	17 / 1

3.2 Equipment and Design

Four clinical measures of color vision were evaluated: (1) RCCT, version 12.1; (2) WCCVT, no version – software date stamped 30 Oct 2012, copyright 2011; (3) CAD, version 2.1.3; and (4) Oculus anomaloscope, version 1.30, type 47715.

The RCCT (Innova Systems, Burr Ridge, IL) is based on the principle of cone isolation [19,20]. Using the staircase mode, a single letter stimulus is presented for 4 seconds on a gray background. The chromaticity of the stimulus and background is selected such that only a single normal cone type (L, M, or S) is sensitive to the target. The subject identifies the stimulus by using a mouse to select the matching letter from a 10-letter answer template. If correct responses are given, the contrast of the stimulus is decreased in a stair-step fashion; similarly, the contrast is sequentially increased with incorrect responses. Step-wise adjustments of contrast for each chromaticity being tested provide an estimate of the just-noticeable contrast for each cone type. Testing is performed at 36 inches under monocular conditions in an otherwise darkened room.

Each cone type is scored on a scale from 0 to 100 in increments of 5, with 75 or greater representing a passing score (additional details concerning the cone contrasts and calibration can be found elsewhere [20]).

The WCCVT (currently marketed as ColorDx by Konan Medical USA, Irvine, CA) resembles a digitized version of pseudo-isochromatic plates and uses targets and background colors that take advantage of confusion lines within the Commission Internationale de l'Eclairage colorspace. A "hidden" image is presented for 2 seconds and then replaced by a template of eight possible answers. Subjects are given unlimited time to identify the number seen in the image or choose that they did not see anything. The first module consists of 25 images with an additional sample image that can be seen by both color normal and color abnormal subjects. Correct identification of at least 21 slides is required for a passing score per Dr. Terry Waggoner, the developer of the test. Subjects who do not meet these minimum criteria are then given two sets of 32 desaturated slides that are intended to delineate between a deutan and protan deficiency, with the diagnosis based on the lower score of the two slide sets. All subjects are also given a set of 12 slides that tests for a tritan deficiency and requires nine or more correct responses to pass. Testing was performed under binocular conditions at a distance of 24-30 inches in an otherwise darkened room.

The CAD (City Occupational Ltd., London) is based on the concept of camouflage wherein a target is presented within a variegated background of randomly varying brightness [21]. During the test, the subject views a colored stimulus moving along the diagonal on a gray background. Both the stimulus and background are formed by discrete elements that dynamically flicker (although the average luminance of both the target and background remains fixed and equal) to produce luminance noise and, thus, limit identification of the motion to color cues alone. The subject identifies the direction of the motion by pressing one of four buttons on a keypad. If the motion cannot be detected, the subject must guess before the test will continue. The contrast of the color stimulus is increased or decreased, depending on whether the response is correct or not, until a threshold is reached. The test screens for normal color vision, quantifies the level of chromatic sensitivity, and diagnoses the type of deficiency (protan, deutan, tritan) when present. Results are reported in standard CAD linear units, which were converted to log units for reporting and correlation with other findings. The pass/fail criteria are proprietary, however, based on prior experience, it appears to be in the range of 1.78 linear units or 0.25 log units. The CAD is performed under binocular conditions at a distance of 1.4 meters in an otherwise darkened room.

The anomaloscope was a computer-controlled Heidelberg Multi-Color (Oculus USA, Arlington, WA) using the Rayleigh equation for red/green color deficiencies. During the test, the subject views a circular color pattern through an eyepiece (much like looking through a microscope). The top half of the pattern is a mixture of red (670 nm) and green (535 nm) light, while the bottom reference field is a single wavelength yellow light (589 nm). After focusing the eyepiece, the subject is shown presentations of varying ratios of red and green in the test pattern and asked if the mixture matches the reference. Once the subject reports an approximate match, he/she is allowed to manually adjust the red/green mixture to provide an exact match (protans are additionally allowed to vary the brightness of the reference light). Once an exact match is achieved, referred to as the "matching midpoint" (MM), the administrator then probes around the match point to determine the range of red/green mixes the subject is willing to call a match. This is referred to as the "matching range" (MR) Throughout this process, the target is presented for 5 seconds with a 1-second presentation of a white

(bleaching) light to minimize color adaptation. Normal color status was defined as an MM between 34.0 and 46.0 with an MR less than 4.1. Midpoints below 34.0 were diagnostic of a deutan deficiency, while midpoints above 46.0 represented a protan deficiency.

The operational task was developed by scientists at the U.S. Air Force School of Aerospace Medicine Operational Based Vision Assessment (OBVA) laboratory with support from U.S. Air Force (USAF) pilot subject matter experts (SMEs). This task simulated the secondary multi-function display of a fifth generation fighter aircraft, or situation awareness (SA) display, which uses colored symbologies to identify "friend" icons (green) versus "foe" icons (red) as well as colored lines to identify various boundary markers (as illustrated in Figure 1). In total, 120 SA display snapshots were created by OBVA researchers in coordination with a pilot SME with over 4,000 hours of flight time. The SA display task was identified as a critical color dependent flight task by a team of vision scientists, aeromedical personnel, and pilot SMEs from several different weapons platforms (the OBVA Integrated Product Team). The chromaticity and luminance of all icons and markers matched those present on the display in the aircraft, and the monitor was calibrated for luminance and chromaticity using a Minolta CS-200.

The subject's task was to determine if a "foe" icon was present between the red and cyan boundary markers. The view shown on the left of Figure 1 represents a positive response given that a set of "foe" icons is located between the specific boundaries, while the view on the right would be a negative response as the "foe" icons are located outside the boundaries. Patient responses were captured using a standard keyboard. A "Yes" response is denoted with the Up arrow and a "No" response with the Down arrow.

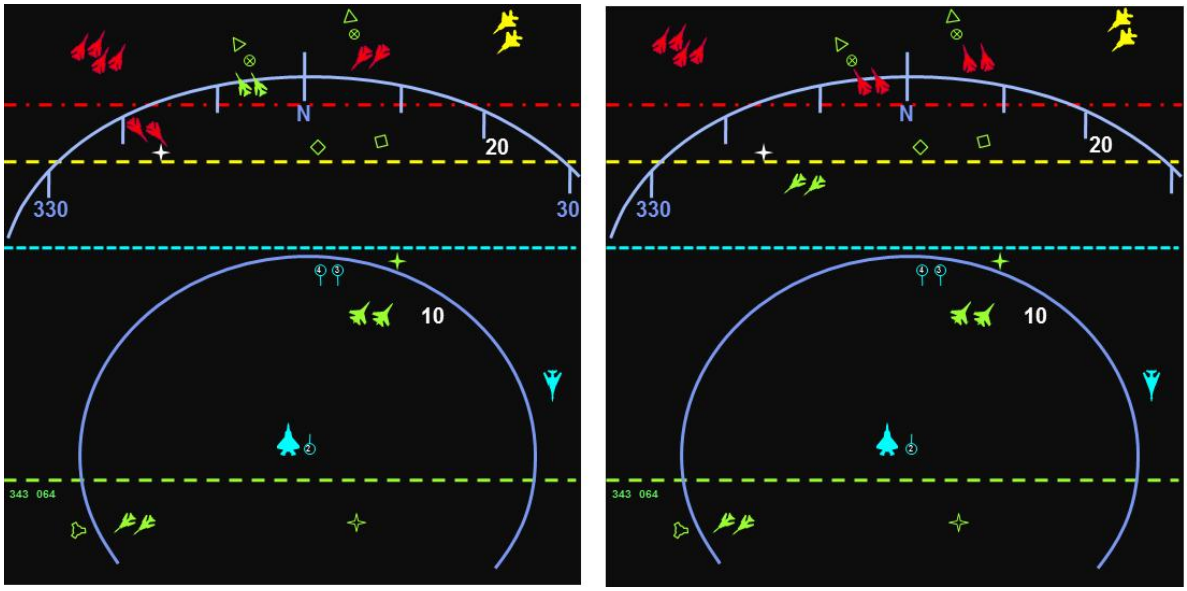

Figure 1. Screenshot of operational task.
View on left represents positive response; view on right represents negative response.

Prior studies utilizing this task have demonstrated a relatively high degree of variability within both color normal and color deficient populations. To control for other variables not related to color vision status such as reaction time, decision making speed, visual tracking skills, etc., the task was initially performed under achromatic (grayscale) conditions where the subject had to identify if a "foe" airplane (denoted by icons of increased luminance) was present in the area of interest (denoted by boundary markers with increased luminance). Results for the chromatic condition were divided by results from the achromatic condition to produce what we refer to as the "normalized" result.

Subjects were given instructions to respond as quickly and accurately as possible with the emphasis on speed. They were allowed to practice both the chromatic and achromatic scenarios until it was evident that there was no additional improvement due to learning effects. Once the test sequence began, each condition (achromatic and chromatic) consisted of three blocks of 50 presentations. Throughput was the performance metric and was calculated as the average percent correct divided by average reaction time for all 150 presentations.

4.0 RESULTS

4.1 Clinical Tasks

Table 2 reports the screening results for each of the four devices. Color vision status for all subjects was based on the consensus of the test battery. Other studies often consider the anomaloscope as the gold standard and assign color status strictly from this device. However, there was compelling evidence that two subjects (#81 and #90) were color anomalous despite being identified as color normal by the Oculus anomaloscope. A third subject (#4) was identified as color abnormal by the Oculus anomaloscope but was clearly normal on all other devices. Table 3 reports findings for these three subjects. Determination of the type of color deficiency for subject #90 was challenging, as the RCCT and WCCVT identified a protan deficiency while the CAD diagnosed a deutan deficiency. For this single subject, the Nagel anomaloscope was used to aid in the diagnosis, and it confirmed a protan deficiency.

Table 2. Screening Characteristics for Each of the Four Clinical Devices Evaluated

Characteristic	RCCT	CAD	WCCVT	Oculus Anomaloscope
Sensitivity (True Positive)	100.0% (50/50)	100.0% (50/50)	94.0%[a] (47/50)	96.0% (48/50)
Specificity (True Negative)	100.0% (50/50)	100.0% (50/50)	94.0% (47/50)	98.0% (49/50)
Type of deficiency correctly diagnosed (i.e., deutan vs. protan)	100.0% (50/50)	94.0% (47/50)	78.0%[b] (39/50)	96.0% (48/50)

[a]17 subjects additionally identified as having a tritan (blue cone) defect, not confirmed on any other testing.
[b]$p<0.05$ versus other devices using Fisher exact test.

Table 3. Subjects Whose Anomaloscope Findings Were Inconsistent with Overall Test Battery

Subject Number	RCCT (OD/OS)	CAD	WCCVT	Oculus Anomaloscope (OD/OS)
4 Color normal by battery, protan by Oculus anomaloscope	L Cone: 95/95 M Cone: 90/95 (Normal)	Red/Green: 1.15 (Normal)	Screening: 23/25 (Normal)	MM: 47.5/48.8 MR: 1.2/0.8 **(Protan)**
81 Color abnormal by battery, normal by Oculus anomaloscope	L Cone: 95/90 M Cone: 60/70 **(Deutan)**	Red/Green: 2.08 **(Deutan)**	Screening: 1/25 Deutan plates: 15/32 Protan plates: 17/32 **(Deutan)**	MM: 37.4/38.1 MR: 0.9/0.9 (Normal)
90[a] Color abnormal by battery, normal by Oculus anomaloscope	L Cone: 60/70 M Cone: 90/95 **(Protan)**	Red/Green: 2.21 **(Deutan)**	Screening: 20/25 Deutan plates: 25/32 Protan plates: 22/32 **(Protan)**	MM: 42.2/39.2 MR: 1.1/3.2 (Normal)

OD = oculus dexter, right eye; OS = oculus sinister, left eye.
[a]Subject #90 determined to be protan on Nagel anomaloscope (MM= 46.0/46.8, MR= 4.0/6.0).

Using the consensus of the test battery to define color status, the RCCT and CAD had the best sensitivity (identifying a color anomalous subject as abnormal) and specificity (identifying a color normal subject as normal) of all devices tested, 100% for both. However, the CAD identified three subjects as deuteranomalous when the abundance of evidence demonstrated a protanomalous deficiency. The WCCVT showed the poorest ability to delineate the nature of the color deficiency, misidentifying the nature of deficiency on 11 of the 50 color anomalous subjects and additionally labelled 17 subjects as having a tritan deficiency that was not observed on any other device and would be a completely unexpected rate of occurrence for this exceptionally rare condition [22].

Figures 2 and 3 report rank-order plots of RCCT and CAD scores. RCCT scores were based on the lower of the L and M cone scores. Figures 4 and 5 report WCCVT scores. Figure 4 shows the screening results for all 100 subjects, while Figure 5 reports results from the desaturated slides correlating to the affected cone for color deficient subjects.

Figure 6 reports rank order of Oculus anomaloscope results for each eye. Vertical lines represent the MR and dashed horizontal lines represent the USAF normal range. Results below the range are consistent with a deutan deficiency and results above represent a protan deficiency. It is clearly apparent that some color abnormal subjects had an MM within normal, but had an extended MR (>4.1 or ± 2.05 units above and below the midpoint). This is considered an abnormal anomaloscope finding and, in this event, color status was defined by the remainder of the test battery.

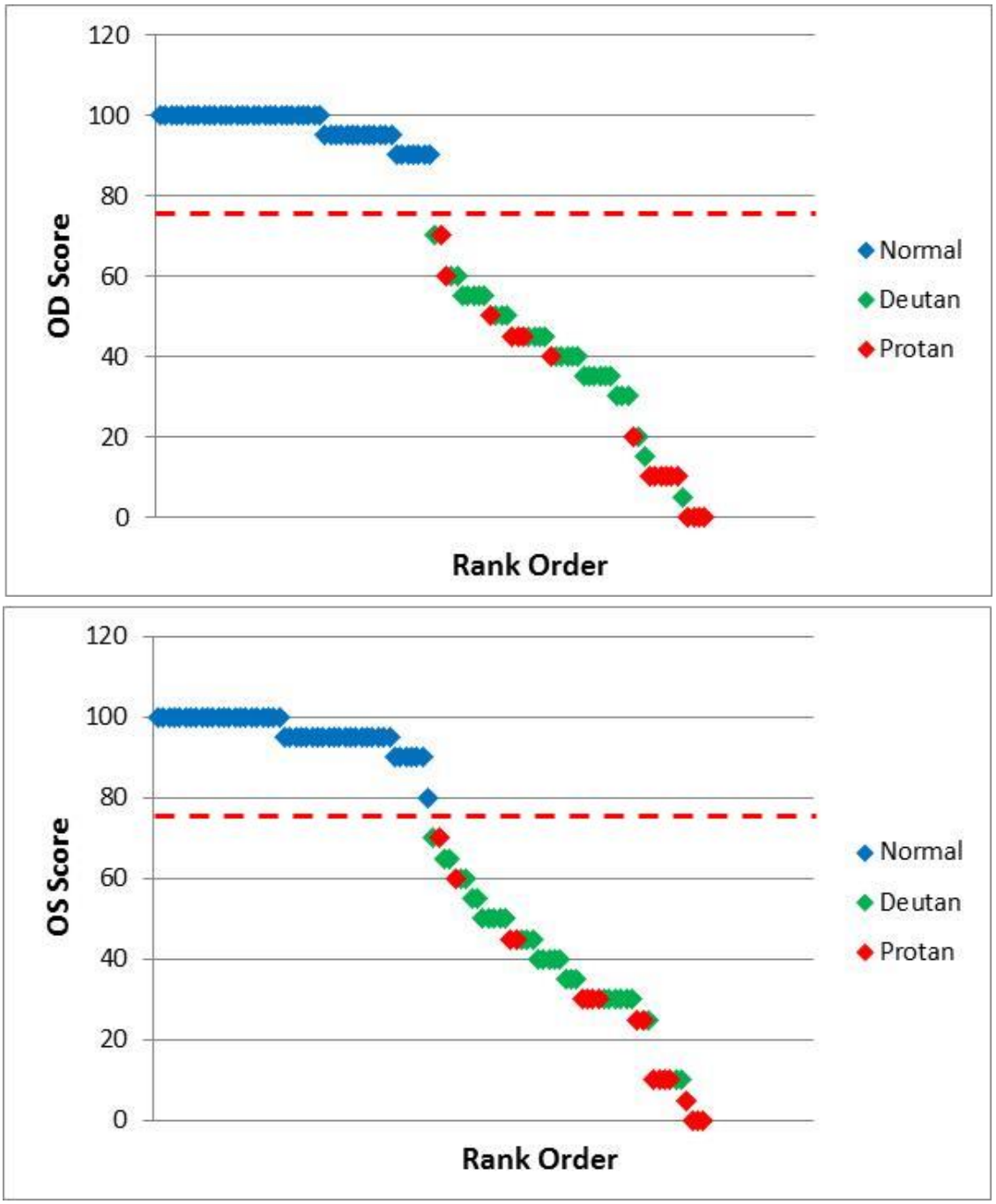

Figure 2. Rank order of RCCT scores. *Dashed line represents USAF pail/fail criteria.*

Figure 3. Rank order of CAD scores. *Pass/fail criteria are proprietary to the manufacturer.*

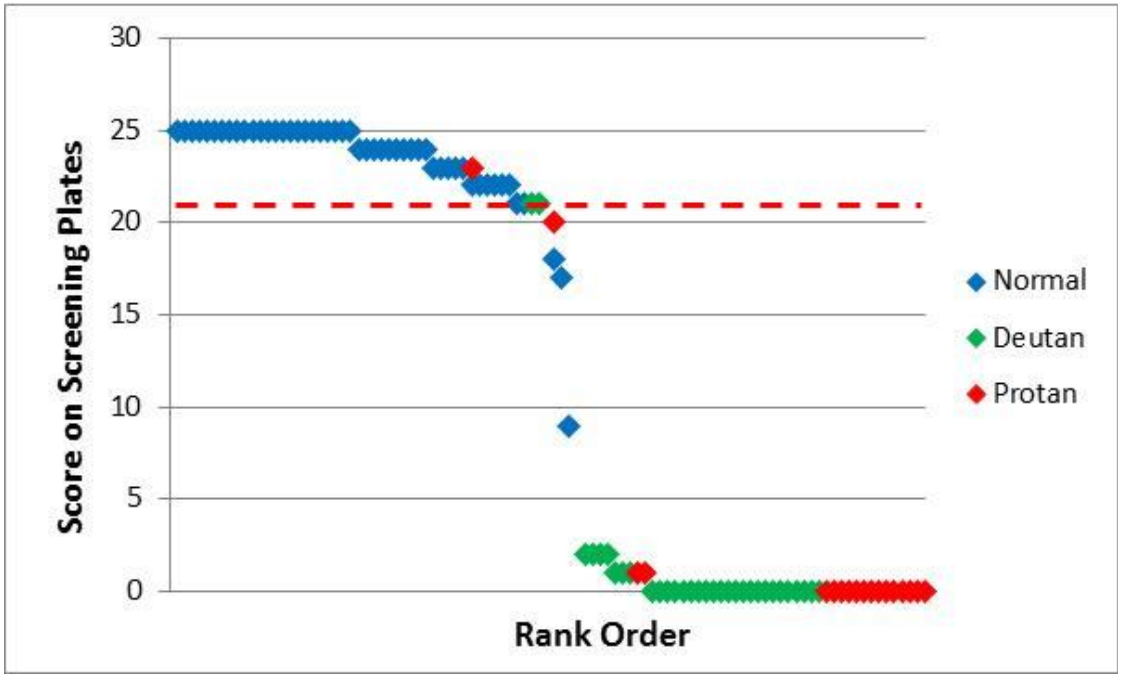

Figure 4. Rank order of WCCVT scores on screening plates.
Dashed line represents pass/fail criteria suggested by Dr. Terry Waggoner.

Figure 5. Rank order of color abnormal scores on WCCT desaturated plates.

Figures 7 and 8 report correlation of RCCT results to the CAD and WCCVT. The RCCT is performed under monocular conditions, while the CAD and WCCVT are administered binocularly. Thus, CCT scores are reported as the average of OD and OS scores. For color deficient subjects, results of the affected cone are reported. For color normal subjects, results of the L and M cone are averaged. Further, the CAD reports result in linear units that were converted to logarithmic units for better comparison to the RCCT. It is readily observed that although there is good correlation between RCCT and WCCVT, it is primarily due to the large pool of normal subjects and associated ceiling effect with each test. As shown in Figure 9, correlation is improved if the desaturated plate score is used for color abnormals.

Figure 10 shows correlation of the RCCT to the anomaloscope. Each test is performed under monocular conditions; however, for the sake of brevity, scores are reported as the average of the right and left eye. The CCT result was the score of the affected cone for color deficient subjects and the average of L and M scores for color normal subjects. The anomaloscope MM was converted from a raw score to the number of standard deviations (SDs) away from the mean MMs of this study's color normal population (40.89 ± 2.11). This effectively negates the direction of the midpoint shift (lower for deutan subjects, higher for protan subjects) and makes comparison to the RCCT more evident. Although not shown graphically, a similar comparison of RCCT scores to the anomaloscope MR was performed and yielded a relatively low correlation, r^2=0.27.

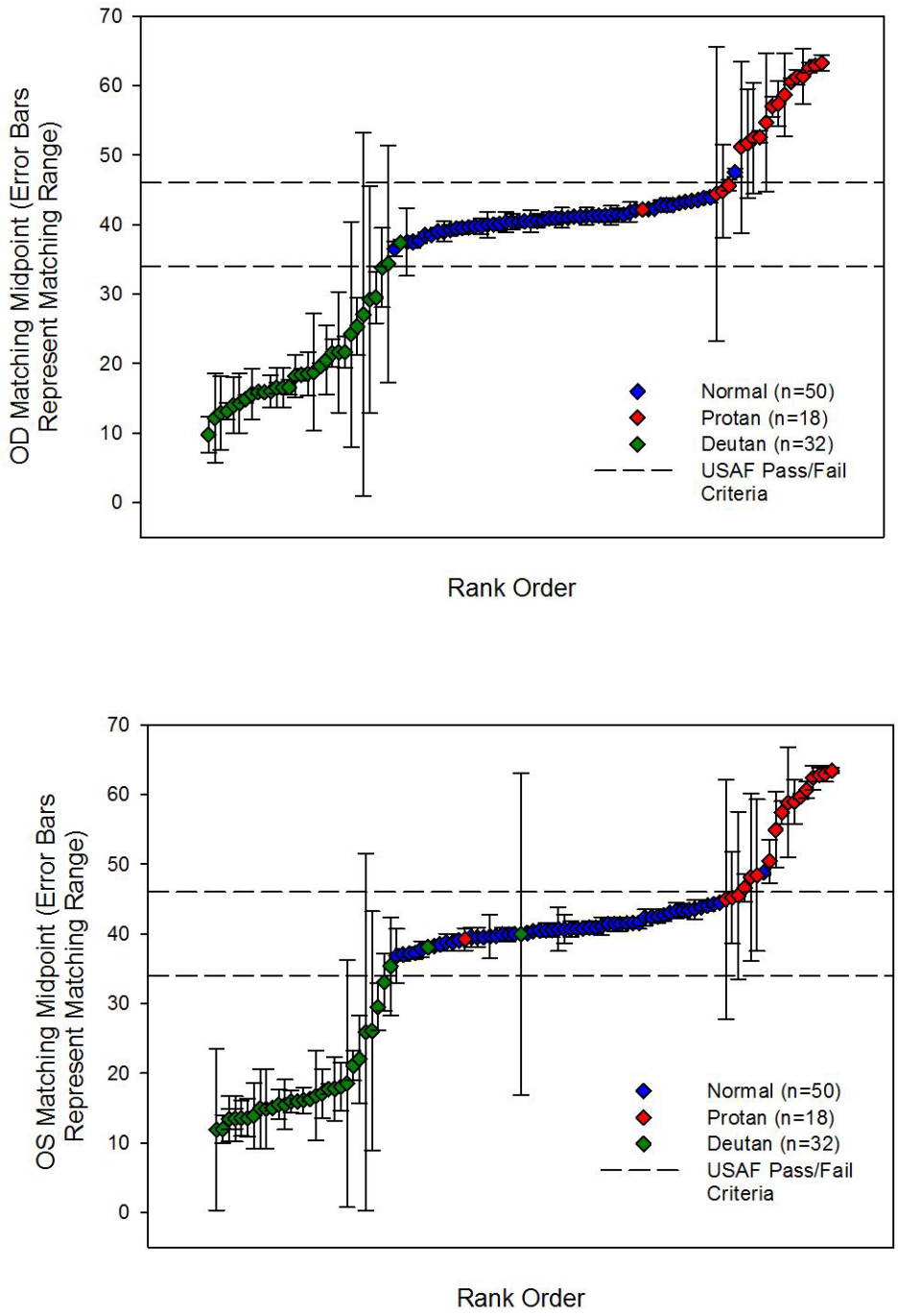

Figure 6. Rank order of Oculus anomaloscope scores.

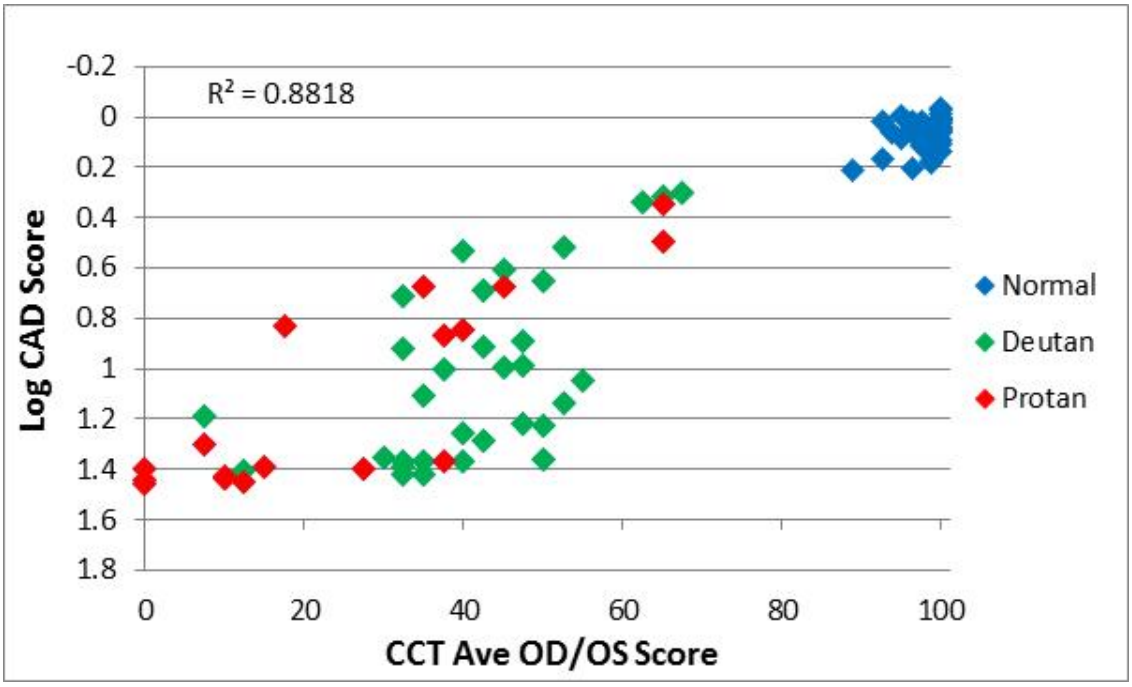

Figure 7. Correlation of RCCT to CAD.

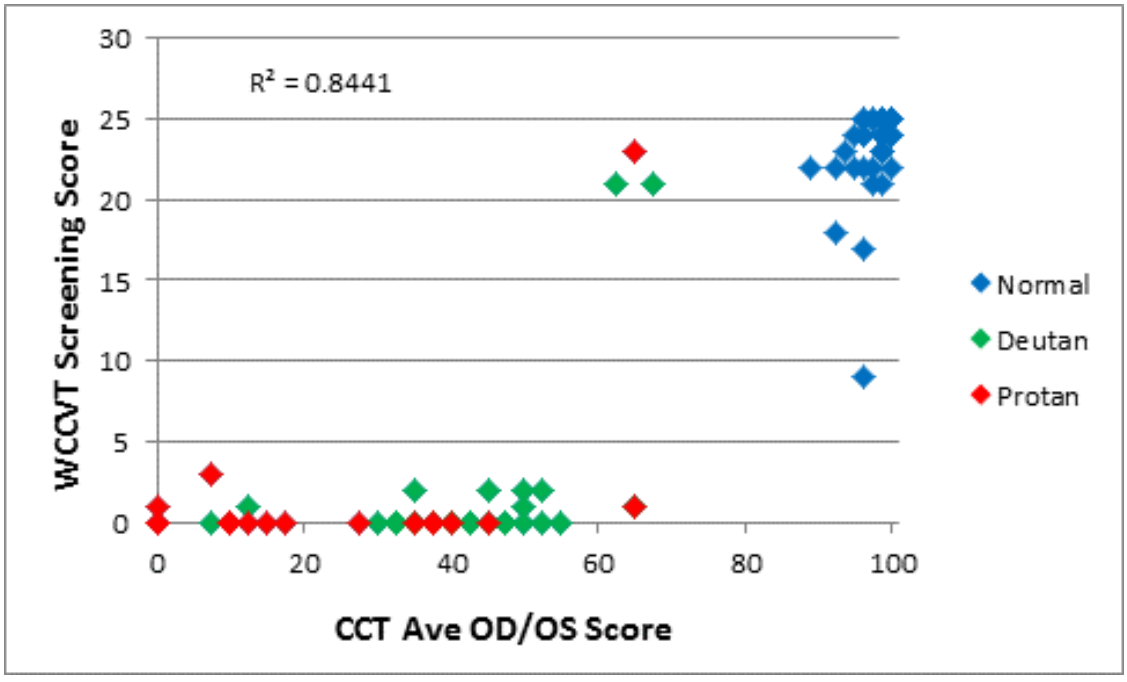

Figure 8. Correlation of RCCT to WCCVT screening scores.

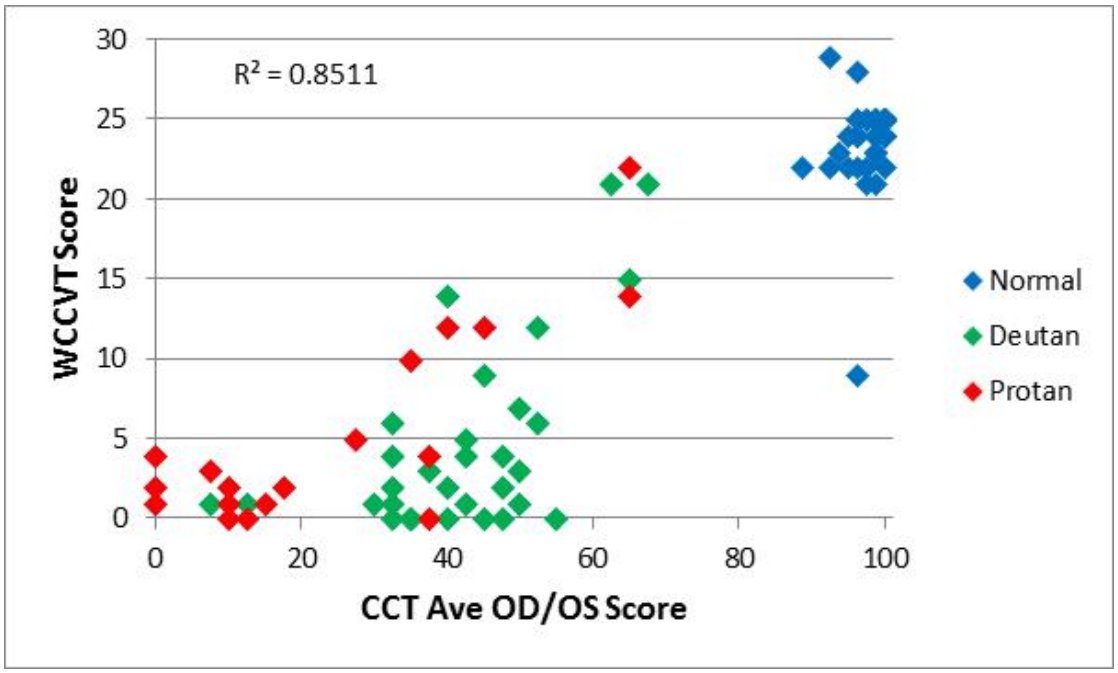

Figure 9. Correlation of RCCT to WCCVT scores from screening plates on normal subjects and scored from desaturated plates corresponding to the affected cone type for color abnormal subjects.

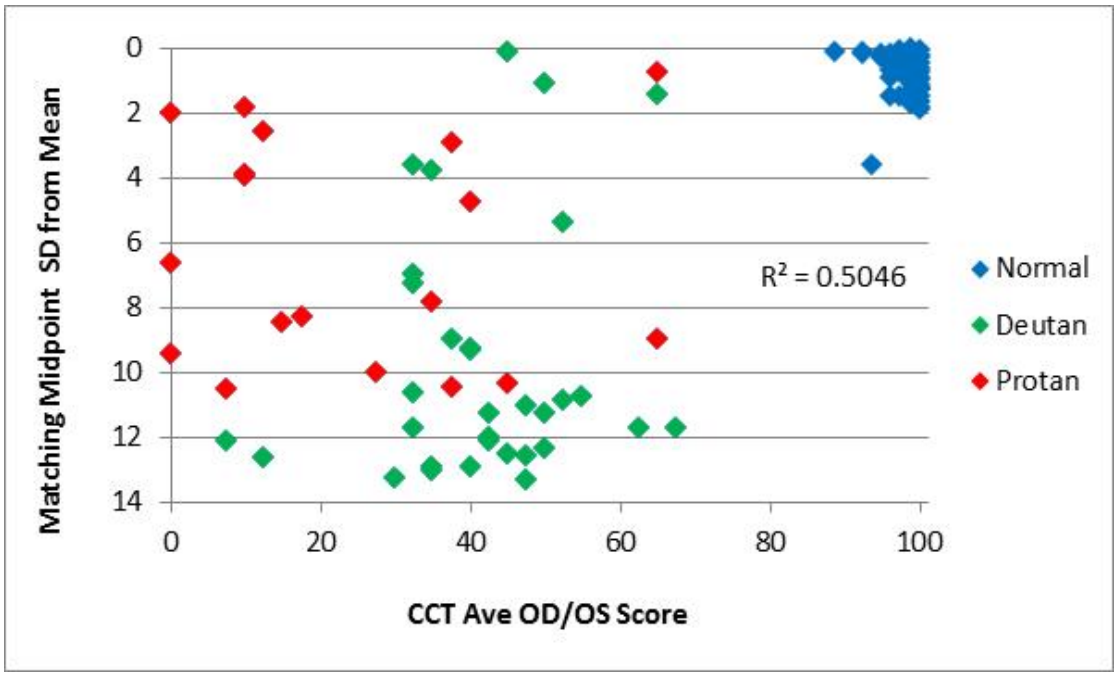

Figure 10. Correlation of RCCT to anomaloscope matching midpoint shift from mean (in SD units).

Table 4 reports correlation coefficients between the three automated tests for all subjects as well as the anomaloscope, while Table 5 lists coefficients only within the abnormal population. RCCT and anomaloscope scores were averaged across the two eyes to allow for comparison with binocular CAD and WCCVT results. Correlations for all devices were significant when all subjects were considered. When the analysis was limited to color abnormal subjects, correlations between automated devices were significant; however, no findings were significant for the anomaloscope.

Table 4. Correlation of Determination between Clinical Devices for All Subjects

Device	RCCT	CAD	WCCVT[a]	WCCVT[b]
CAD	0.882			
WCCVT[a]	0.844	0.812		
WCCVT[b]	0.851	0.894	0.914	
Anomaloscope[c]	0.507	0.458	0.593	0.579

Note: $p<0.001$ for all results.
[a]Using screening plates only.
[b]Using desaturated plates for color abnormals.
[c]MM shift from mean level in units of SDs.

Table 5. Correlation of Determination between Clinical Devices for Color Abnormal Subjects

Device	RCCT	CAD	WCCVT[a]	WCCVT[b]
CAD	0.446[c]			
WCCVT[a]	0.173[c]	0.256[c]		
WCCVT[b]	0.294[c]	0.639[c]	0.570[c]	
Anomaloscope[d]	0.016	0.058	0.003	0.014

[a]Using screening plates only.
[b]Using desaturated plates for color abnormals.
[c]$p<0.002$.
[d]MM shift from mean level in units of SDs.

4.2 Operational Task

Table 6 reports correlations of the clinical tests to the operational task. There was modest correlation between each device and performance on the task under chromatic conditions and no correlation between the clinical tests and achromatic findings. Better correlation was achieved for each device when the throughput was normalized (i.e., chromatic throughput/achromatic throughput). The CAD yielded the highest level of correlation with the operational task (Figures 11-13), while the anomaloscope had the lowest.

Table 6. Correlation of Determination between Clinical Screenings and Throughput on Operational Task

Device	Chromatic Task (all p<0.0001)	Achromatic Task (no significance)	Ratio of Chromatic to Achromatic (all p<0.0001)
RCCT	0.276	0.005	0.484
CAD	0.324	0.001	0.514
WCCVT[a]	0.255	0.007	0.464
WCCVT[b]	0.282	0.006	0.496
Anomaloscope[c]	0.164	0.014	0.342

[a]Using screening plates only.
[b]Using desaturated plates for color abnormals.
[c]MM shift from mean level in units of SDs.

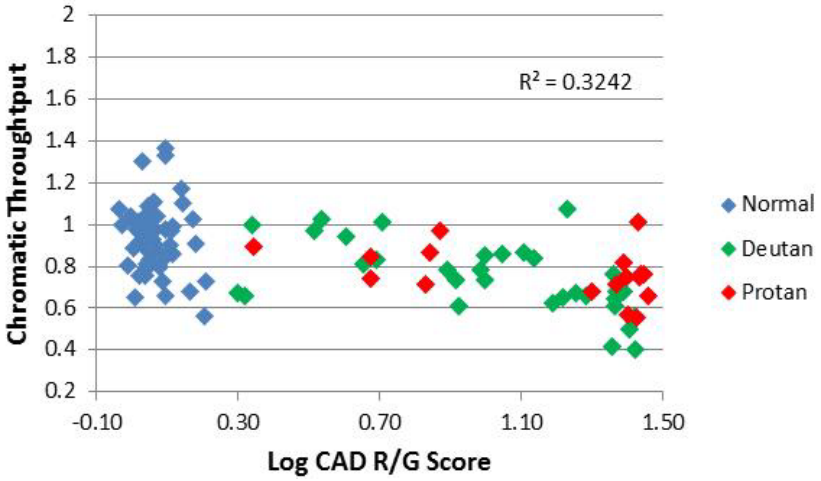

Figure 11. Chromatic throughput as a function of CAD scores.

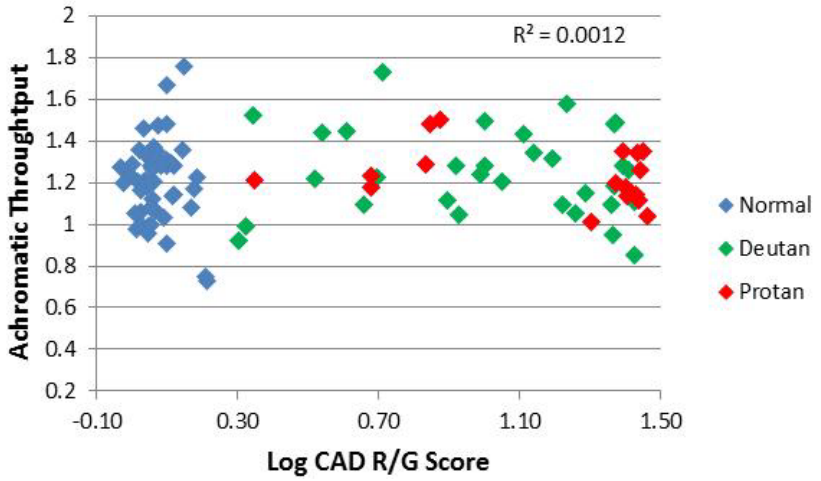

Figure 12. Achromatic throughput as a function of CAD scores.

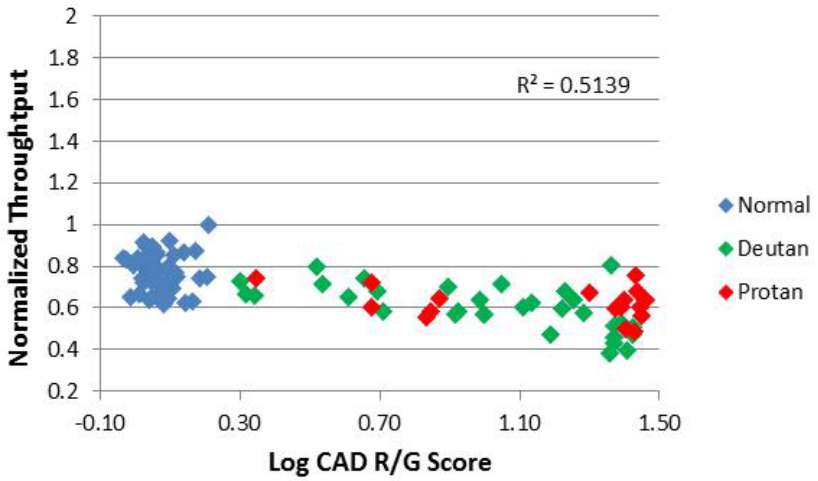

Figure 13. Normalized throughput (chromatic/achromatic) as a function of CAD scores.

Figure 14 shows performance on the operational task based on color status. Subjects were placed into one of four groups: (1) color normal; (2) mildly deficient, defined as an average OD/OS RCCT score of 55-70 on the affected cone type; (3) moderately deficient, defined as an average RCCT score of 35-50; and (4) severely deficient, defined as an average RCCT score less than 35. These groups were selected based on the current grading scale used by the USAF for defining color vision status. Subjects were not separated into deutan or protan specific subgroups, as the sample size in the mildly deficient group would have been inadequate to derive meaningful results. Statistical differences in performance between the four groups on the three aspects of the operational task were analyzed using a chi-square test and are reported in Tables 7-9. Under chromatic conditions, moderate and severe color deficients performed (statistically) significantly worse than color normals, while no significant differences were observed between mild deficients and color normals. There was, additionally, a significant difference between moderate and severe color deficients, but not between mild and severe deficients, likely due to the limited sample size in the mild color deficient group. No significant differences in performance between any group were observed under achromatic conditions. When chromatic performance was normalized by achromatic performance, both moderate and severe deficients performed statistically worse than normals. Severely deficient subjects also performed statistically worse than both mild and moderate deficients.

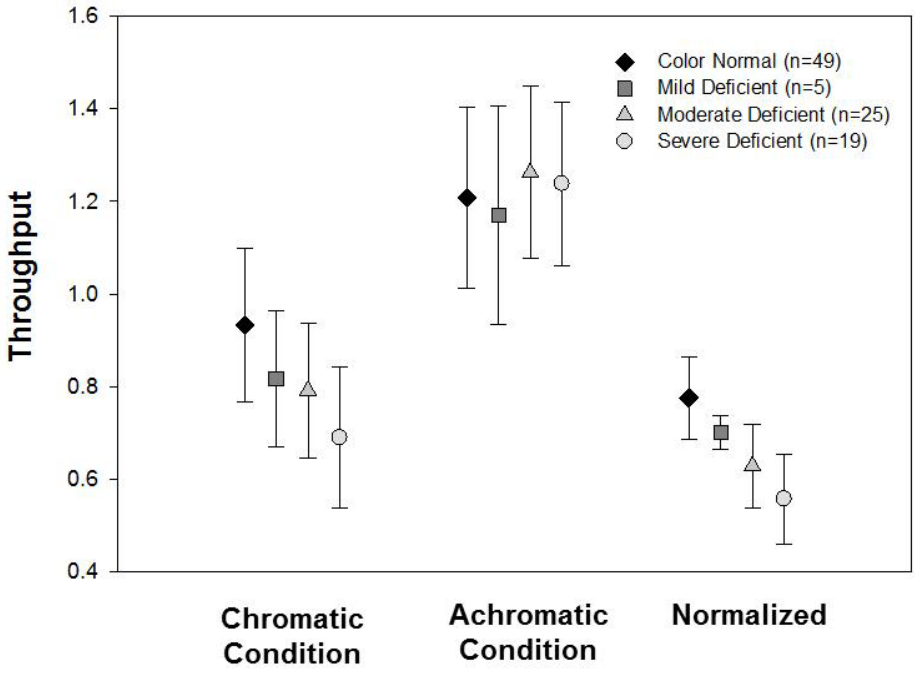

Figure 14. Relative performance on operational task based on color status (mean ± 1 SD).

Table 7. Effect of Color Vision Status on Chromatic Throughput on Operational Task (p-values for chi-square Statistic)

Status	Color Normal	Mild Deficient	Moderate Deficient
Mild Deficient	0.142		
Moderate Deficient	0.001	0.728	
Severe Deficient	<0.001	0.109	0.030

Table 8. Effect of Color Vision Status on Achromatic Throughput on Operational Task (p-values)

Status	Color Normal	Mild Deficient	Moderate Deficient
Mild Deficient	0.683		
Moderate Deficient	0.259	0.339	
Severe Deficient	0.561	0.476	0.669

Table 9. Effect of Color Vision Status on Normalized Throughput (Chromatic/Achromatic) on Operational Task (p-values)

Status	Color Normal	Mild Deficient	Moderate Deficient
Mild Deficient	0.073		
Moderate Deficient	<0.001	0.092	
Severe Deficient	<0.001	0.004	0.016

5.0 DISCUSSION

All of the computer-based tests evaluated (RCCT, CAD, and WCCVT) proved to be highly effective in screening for color vision deficiencies. Overall, there was good agreement between the computer-based tests and the anomaloscope in classifying observers as color normal or abnormal, although there were three subjects where discrepancies were observed. One may conclude that the computer-based tests were wrong, or alternatively, that the anomaloscope was in error. We would propose that neither of these conclusions is true. Rather, the anomaloscope is measuring a different visual attribute than the computer-based tests, and when color deficiency is small, classification by the anomaloscope may conflict with other tests. The interest of the USAF is to screen aircrew applicants and identify those who possess qualities that give them the best chance to succeed in the aviation environment. This study demonstrates that, of all of the clinical devices evaluated, the anomaloscope correlated with operational performance the poorest. Further, correlations between the anomaloscope and the automated tests were much poorer than correlations between the three computer-based tests. Thus, while the anomaloscope may continue to be considered as the "gold standard" for defining color vision deficiency, the CAD, CCT, and CCVT are more suitable for aeromedical color vision screening.

Overall, there was good agreement between the computer-based tests. Coefficients of determination (R^2) were over 0.80 between devices. When the analysis was limited to color abnormal subjects, correlations were more modest, ranging from 0.173 to 0.639. Correlations between the computer tests and the anomaloscope were modest for all subjects, and there was no significant correlation when the analysis was limited to color deficient subjects.

Findings from the operational task proved to be highly variable among both color normal and color anomalous subjects, with only ~30% of the variance attributable to color vision status. The task was replicated under grayscale conditions as an attempt to tease out other, non-color vision variables that would contribute to overall performance. When the throughput under chromatic conditions was normalized by dividing by the throughput under achromatic conditions, the coefficient of determination values increased almost twofold.

The operational task demonstrated a general trend of poorer performance with worsening levels of color deficiency. Moderate and severely color deficient subjects performed at a level statistically worse than normals under both the chromatic condition as well as the normalized condition. In contrast, mildly deficient subjects showed no statistical differences from color normal subjects under any condition, although this finding should be taken with caution as the sample size for mild color deficient subjects was low (n=5 subjects). No differences were seen between any groups on the achromatic test.

6.0 CONCLUSIONS

Electronic measures of color vision screening demonstrated high sensitivity and specificity; however, this was a relatively small study and these results should be validated with a much larger sample size.

The operational task demonstrated that subjects with color vision deficiencies were poorer at processing color-coded information; however, many color deficient subjects, including some with moderate and severe deficiencies, were able to perform at a level comparable to color normal subjects. Furthermore, while this effect was statistically significant, the magnitude was small. This finding should be interpreted carefully. Our study used colors with very specific chromaticities and with high contrast. Testing was performed under optimal conditions on a well-calibrated display. True operational environments are far more austere, with variables that go beyond those evaluated in this study. Further studies with different color chromaticities, contrast levels, and operational scenarios are indicated.

Although the anomaloscope has been considered the "gold standard" for defining color status, its results did not correlate with other tests in terms of the severity of a color vision anomaly and correlated poorly with performance on the operational task. The RCCT, CAD, and WCCVT produced more meaningful results in this regard.

7.0 REFERENCES

1. National Transportation Safety Board. Near head-on collision and derailment of two New Jersey Transit commuter trains near Secaucus, New Jersey, February 9, 1996. Washington (DC): NTSB; 1997. Railroad Accident Report PB97-916301; NTSB/RAR-97/01. [Accessed 30 Sep 2014]. Available from http://www.ntsb.gov/investigations/AccidentReports/Pages/RAR9701.aspx.
2. National Transportation Safety Board. Head-on collision of two Union Pacific Railroad freight trains near Goodwell, Oklahoma, June 24, 2012. Washington (DC): NTSB; 2013. Accident Report NTSB/RAR-13/02; PB2013-107679. [Accessed 30 Sep 2014]. Available from http://www.ntsb.gov/investigations/AccidentReports/Pages/RAR1302.aspx.
3. National Transportation Safety Board. Collision with trees on final approach, Federal Express flight 1478, Boeing 727-232, N497FE, Tallahassee, Florida, July 26, 2002. Washington (DC): NTSB; 2004. Aircraft Accident Report NTSB/AAR-04/02; PB2004-910402, Notation 7501B. [Accessed 30 Sep 2014]. Available from http://www.ntsb.gov/investigations/AccidentReports/Pages/AAR0402.aspx.
4. Wilson G. Researches on colour-blindness. Edinburgh: Sutherland & Knox, South Bridge; Simpkin, Marshall and Co., London; 1855. [Accessed 1 Oct 2014]. Available from http://books.google.com/books?id=_YN0Ve_jXBUC&pg=PR11&source=gbs_selected_pages&cad=2#v=onepage&q&f=false.
5. Jeffries BJ. Color-blindness: its dangers and its detection. Boston (MA): Houghton, Osgood and Company; 1880:261-263,289. [Accessed 1 Oct 2014]. Available from https://archive.org/ stream/colorblindnessit00jeff#page/n7/mode/2up.
6. Vingrys AJ, Cole BL. Origins of colour vision standards within the transport industry. Ophthalmic Physiol Opt. 1986; 6(4):369-375.
7. Mollon JD, Cavonius LR. The Lagerlunda collision and the introduction of color vision testing. Surv Ophthalmol. 2012; 57(2):178-194.

8. Jennings JE. A self-recording device for testing the color sense. JAMA. 1914; LXIII(12):1030-1031.
9. Wilmer WH, Berens C Jr. Department of ophthalmology. In: Wilmer WH. Aviation medicine in the A.E.F. Washington (DC): Government Printing Office; 1920:172. [Accessed 3 Oct 2014]. Available from http://babel.hathitrust.org/cgi/pt?id=uc1.%24b124035#view=1up;seq=1.
10. Tredici TJ, Mims JL, Culver JF. History, rationale and verification of colour vision standards and testing in the United States Air Force. AGARD Conference Proceedings No. 99 on Colour Vision Requirements in Different Operational Roles. 1972 May 30; Brussels, Belgium. Neuilly-sur-Seine (France): NATO; 1972. AGARD-CP-99.
11. Monlux DJ, Finne HA, Stephens MB. Color blindness and military fitness for duty: a new look at old standards. Mil Med. 2010; 175(2):84-85.
12. Birch J. Efficiency of the Ishihara test for identifying red-green colour deficiency. Ophthalmic Physiol Opt. 1997; 17(5):403-408.
13. Miyahara E. Errors reading the Ishihara pseudoisochromatic plates made by observers with normal colour vision. Clin Exp Optom. 2008; 91(2):161-165.
14. Regan BC, Reffin JP, Mollon JD. Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. Vision Res. 1994; 34(10):1279-1299.
15. Reffin JP, Astell S, Mollon JD. Trials of a computer-controlled colour vision test that preserves the advantages of pseudoisochromatic plates. In: Drum B, Moreland JD, Serra A, eds. Colour vision deficiencies X. Proceedings of the 10th Symposium of the International Research Group on Colour Vision Deficiencies; 1989 Jun 25-28; Cagliari, Italy. Documenta Ophthalmologica Proceedings Series. 1991; 54:69-76. Berlin: Kluwer Academic Publishers.
16. Barbur J, Rodriguez-Carmona M, Evans S, Milburn N. Minimum colour vision requirements for professional flight crew: recommendations for new colour vision standards. London (UK): Civil Aviation Authority; 2009. CAA Paper 2009/04. [Accessed 14 Oct 2014]. Available from http://publicapps.caa.co.uk/modalapplication.aspx?catid=1&pagetype=65&appid=11&mode=detail&id=3560.
17. Picken D, Mann W, Rings M. Preliminary validation of a computerized color vision test. Poster presented at the 84th Annual Scientific Meeting of the Aerospace Medical Association; 2013 May 12-16; Chicago, IL. [Accessed 14 Oct 2014]. Available from http://www.testingcolorvision.com/navyvalidation.pdf.
18. U.S. Air Force. Medical examinations and standards. Washington (DC): Department of the Air Force; 2015. Air Force Instruction 48-123.
19. Rabin J, Gooch J, Ivan D. Rapid quantification of color vision: the cone contrast test. Invest Ophthalmol Vis Sci. 2011; 52(2):816-820.
20. Rabin J. Quantification of color vision with cone contrast sensitivity. Vis Neurosci. 2004; 21(3):483-485.
21. Safety Regulation Group. Minimum colour vision requirements for professional flight crew - Part 1. The use of colour signals and the assessment of colour vision requirements in aviation. London (UK): Civil Aviation Authority; 2006. Paper 2006/04. [Accessed 14 Oct 2014]. Available from http://publicapps.caa.co.uk/modalapplication.aspx?catid=1&pagetype=65&appid=11&mode=detail&id=2407.
22. Birch J. A practical guide for colour-vision examination: report of the Standardization Committee of the International Research Group on Colour-Vision Deficiencies. Ophthalmic Physiol Opt. 1985; 5(3):265-285.

LIST OF ABBREVIATIONS AND ACRONYMS

CAD	Colour Assessment & Diagnosis
MM	matching midpoint
MR	matching range
OBVA	Operational Based Vision Assessment
RCCT	Rabin Cone Contrast Test
SA	situation awareness
SD	standard deviation
SME	subject matter expert
USAF	U.S. Air Force
WCCVT	Waggoner Computerized Color Vision Test